Data Science for Cyber-Security 2017

Nick Heard, Niall Adams, Patrick Rubin-Delanchy, Melissa Turcotte

Department of Mathematics, Imperial College London
School of Mathematics, University of Bristol
Los Alamos National Laboratory
http://statisticalcyber.com

25 September, 2017

Imperial College London

Why Statistical Cyber-Security?

Statistical Cyber Research

Data science techniques have an important role to play in the next generation of cyber-security defences.

Inside a typical enterprise computer network, a number of high-volume data sources are available which could enable the discovery and prevention of cyber-attacks and other nefarious network activity.

At Imperial, our interests are in developing statistical, probability model-based techniques for identifying the most subtle intrusion attempts using these data sources.

The advantage of such approaches is the ability to pre-learn, from historical data, complex patterns of normal computer and network behaviour, so that anomalies can be detected which would not stand out otherwise; one example is unusual network traversal using legitimate credentials.

Data sources and platforms

- Network flow data — IP→IP, protocol, ports, TCP flags, packets, bytes, time, duration
- Authentication events — usernames, computers, success, type, time
- Host-sensor data — network events, processes, memory usage, time, duration, lock/unlock
- Physical — building access control, sensors, IoT

High volume, high frequency data which require thinning (screening, triage) and parallel processing:

- Hadoop — MapReduce and Spark
- Algorithms which scale well or can run in the stream

Methodological Approaches

From different levels of resolution:

- Entire network analysis - graph theory, spectral decompositions, community detection, clustering. Also high level traffic summaries for network oversight.
- Node-based models - building statistical models of the processes run by a host, its network connectivity, pattern of life.
- Edge-based models - detecting beacons to specific IP addresses, temporal dependence on neighbouring edges, typical packet sizes.

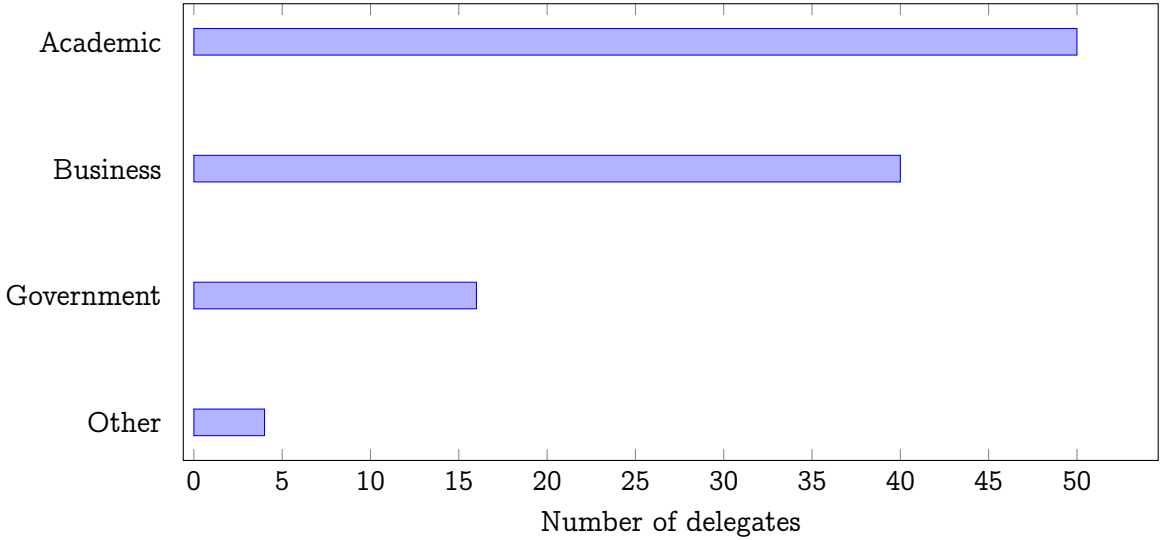
All of these viewpoints, and others, can contribute to the end goal of better cyber-security.
There will not be one statistical test that answers all questions.
Power is to be obtained by combining several possibly weak indicators into a strong overall signal.

Who is here?

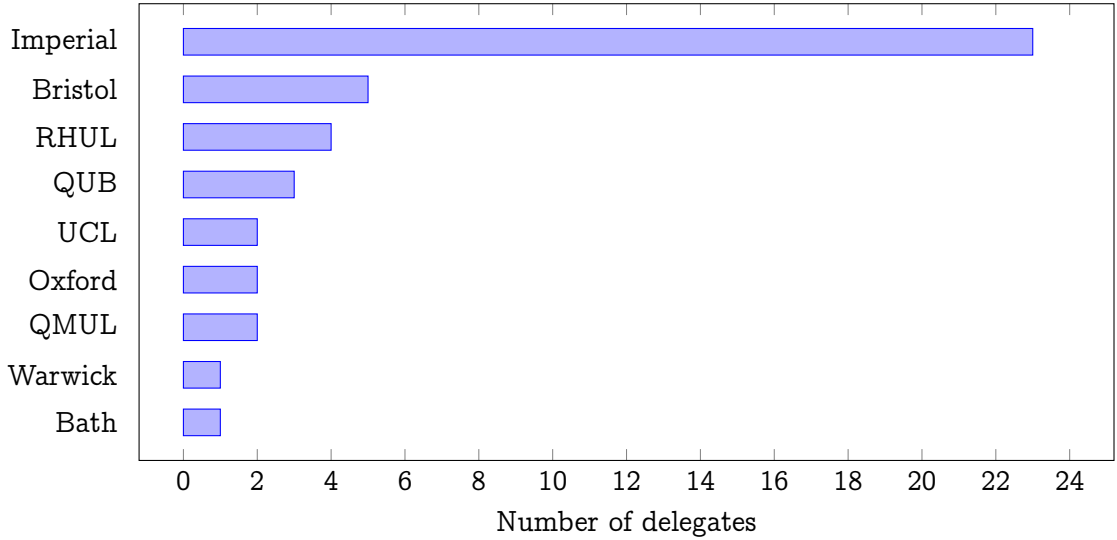
Institutions

ATI
BSI
BT
Barclays
Carnegie Mellon
Cisco Systems
Crossword Cybersecurity
CyberOwl
Dartmouth
Data Done Right
DSTL
ECIT / CSIT
Easy Solutions Inc
Ernst & Young
Facebook
G-Research
GSK
Glasgow School of Art
HMG
HP Inc
Heilbronn Institute
IDA CCS
Imperial College London
Kindred
Los Alamos National Lab
Mentat
NCSC
Palo Alto Networks
QA LTD
Queen Mary, U. of London
Queen's University Belfast
RMS
RSA
Roke Manor Research Ltd
Royal Holloway
Shell
TTP Plc
Thales
US Air Force
US Naval Surface
University College London
University of Bath
University of Bristol
University of Oslo
University of Oxford
University of Warwick
Ydrogios Insurance
Zynga

Sectors


Academic
Business
Government
Other
0
5
10
15
20
25
30
35
40
45
50
Number of delegates

UK Universities

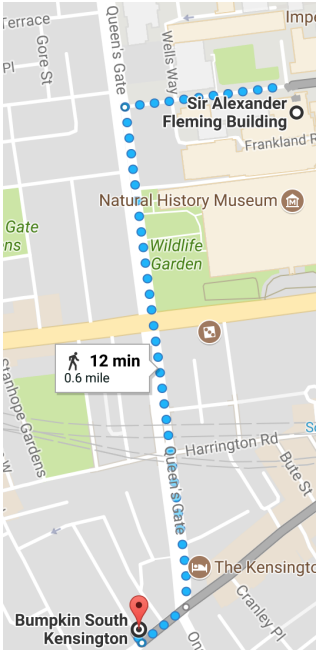


Arrangements

* All presentations in this lecture theatre (G34)
* Teas/coffees and lunches served in the rooms opposite (Rooms 120—122)
* Poster session this evening also in Rooms 120—122
* Data Science Institute Observatory tours:
  - Tues 17:00 https://goo.gl/bk2gB7
  - Tues 17:30 https://goo.gl/NVdqGM
* Dinner tomorrow night, 18:30 @Bumpkin

---

* Full schedule http://statisticalcyber.com



Wi-Fi

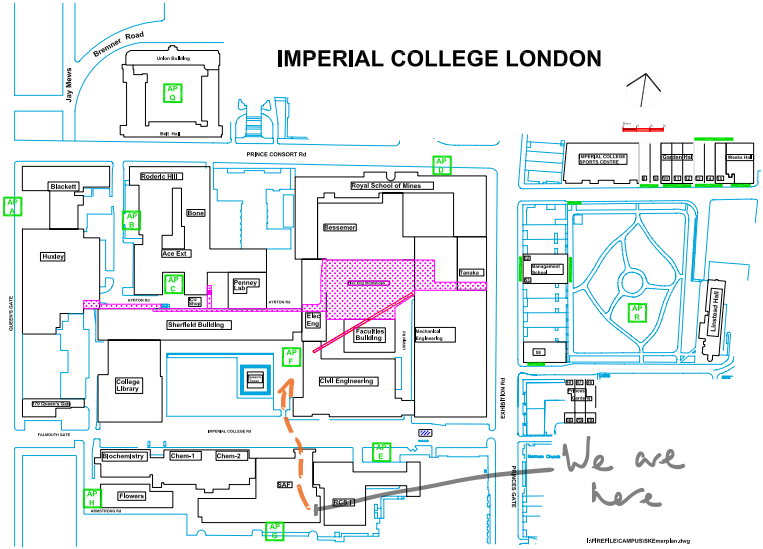
Research and education community:

- eduroam

Everyone else:

- Sky WiFi. Connect to **The Cloud** and follow the instructions to register

Fire Safety


IMPERIAL COLLEGE LONDON
Bremner Road
Jay Mews
Union Building
PRINCE CONSORT Rd
Royal School of Mines
Blackett
Roderic Hill
Huxley
Ace Ext
Bone
Bessemer
Sherfield Building
Faculties Building
Mechanical Engineering
Chem Eng
Civil Engineering
College Library
Biochemistry
Chem-1
Chem-2
Flowers
SAC
SAF
QUEEN'S GATE
EXHIBITION RD
IMPERIAL COLLEGE RD
PALMERSTON GATE
PRINCES GATE
QUEENS GATE
We are here
I:\FIRE\FLEICAMPUS\SKEmerplan.dwg

Thank you to our sponsors

Sponsorship for this workshop has kindly been provided by

- The Heilbronn Institute for Mathematical Research
- Imperial College London
- GSK
- Mentat Innovations
- World Scientific